

SNSN-323-63
November 27, 2018

Observation of $t\bar{t}H$ production with ATLAS

FABRICE HUBAUT,
ON BEHALF OF THE ATLAS COLLABORATION¹

*Centre de Physique des Particules de Marseille,
Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France,
163 Av. de Luminy, Case 902, 13288 Marseille Cedex 9, France*

An overview of the analyses that led to the observation of the Higgs boson production in association with a top quark pair ($t\bar{t}H$) with the ATLAS experiment is given. Up to 80 fb^{-1} of proton–proton collision data recorded by the ATLAS experiment at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ at the Large Hadron Collider (LHC) have been used to achieve this major milestone.

PRESENTED AT

11th International Workshop on Top Quark Physics
Bad Neuenahr, Germany, September 16–21, 2018

¹Copyright 2018 CERN for the benefit of the ATLAS Collaboration. Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

1 Introduction

All measurements of Higgs boson properties are consistent with the Standard Model (SM) predictions, but the fermion sector was so far less tested. In the SM, couplings between the Higgs boson and fermions occur through Yukawa interactions, whose strength is proportional to the fermion mass. The top quark Yukawa coupling is therefore predicted to be by far the largest one, with a value close to unity. Indirect constraints on this coupling can be obtained via the Higgs boson production by gluon-gluon fusion or its decay to two photons, which proceed through loops. However, New Physics (NP) could also enter in these loops, hence a direct measurement is needed to disentangle and probe NP. Such a direct measurement can be achieved through the associated production of the Higgs boson and a top quark pair ($t\bar{t}H$), which is a tree-level process and whose cross-section is proportional to the square of the top quark Yukawa coupling. Observing the $t\bar{t}H$ production would therefore allow to simultaneously constrain this coupling and possible Beyond the Standard Model (BSM) effects.

The $t\bar{t}H$ production is quite a rare process, with a cross-section at the Large Hadron Collider (LHC) of ~ 0.5 pb at a center-of-mass energy of $\sqrt{s} = 13$ TeV. This is typically 1% of the inclusive Higgs boson production cross-section and less than 0.1% of the $t\bar{t}$ production one. This rare production has to be convolved with all possible combinations of $t\bar{t}$ and Higgs boson decays, giving rise to a very rich spectrum of possible signatures. A broad spectrum of analyses has therefore been designed to cover these multiple final states, with their selections designed to avoid any overlap between them. They will be described in the next sections together with their combination, starting from the most abundant ones ($t\bar{t}H$, $H \rightarrow b\bar{b}$ and $t\bar{t}H$, $H \rightarrow (WW^*, \tau\tau, ZZ^*) \rightarrow \text{leptons}$), which use 36.1 fb^{-1} of proton-proton collision data recorded by the ATLAS experiment [1] at a center-of-mass energy of $\sqrt{s} = 13$ TeV, to the more rare ones ($t\bar{t}H$, $H \rightarrow \gamma\gamma$ and $t\bar{t}H$, $H \rightarrow ZZ^* \rightarrow 4l$ ($l = e, \mu$)), which use 79.8 fb^{-1} of data.

2 Measurement in the $H \rightarrow b\bar{b}$ channel

The search for the $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ channel [2] benefits from a large branching ratio, but suffers from a huge combinatoric background at the event reconstruction level and a large irreducible background (mostly $t\bar{t}$ + heavy flavour jets) associated with big theoretical uncertainties. The modelling of this background and the definition of the associated uncertainty scheme are performed using state-of-the-art Monte Carlo (MC) generators and constraints from data in background-enriched regions. The nominal $t\bar{t}$ MC sample uses an inclusive next-to-leading order (NLO) generator normalised to predictions at next-to-next-to-leading order in perturbative

QCD including resummation of next-to-next-to-leading logarithmic soft gluon terms. The main associated systematic uncertainties are based on comparisons against alternative MC samples, using different generators, parton shower or hadronisation models. The normalisations of $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds are derived from data. To define the analysis regions, a selection of semi- and di-leptonic $t\bar{t}$ decays is first performed, based on the presence of one or two light leptons, followed by a further categorisation based on the jet multiplicity and on the value of the b -tagging discriminant of each jet (four b -tagging working points are used). This allows to define 9 signal-enriched regions (SR) and 10 background-enriched control regions (CR), with very different signal purities and background fractions (a “boosted” category targeting high p_T top/Higgs candidates is also defined). The best signal purity in these regions is only $\sim 5\%$, so a cascade of Multivariate Analysis (MVA) classifiers is used to further separate the signal from the backgrounds. A $t\bar{t}H$ system reconstruction Boosted Decision Tree (BDT) is designed in most SR to solve the combinatorics to identify the best parton-jet assignments and reconstruct the final state. It is complemented with a Likelihood discriminant in the single-lepton SR, built from all different jet permutations, and a Matrix Element Method in the single-lepton SR with the best purity. The outputs of all these classifiers, together with the related Higgs/top candidate properties, event kinematic variables and b -tag information, are input into a classification BDT built in each SR to best separate the signal from the backgrounds. A combined profile likelihood fit is performed across all analysis regions to extract the signal yield. The significance of the observed (expected) signal is 1.4σ (1.6σ). The measured signal strength is in good agreement with the SM prediction within the relative measurement precision of 75%, as seen in Figure 1 (left). This precision is dominated by the systematic uncertainties, in particular those associated with $t\bar{t}+\geq 1b$ modelling.

3 Measurement in multilepton final states

The search for $t\bar{t}H$ production in multilepton final states [3] targets $H \rightarrow WW^*$, $H \rightarrow \tau\tau$ and $H \rightarrow ZZ^*$ decays. The $t\bar{t}$ background is suppressed by requiring two same-sign or at least 3 leptons. Seven orthogonal channels, categorised by the multiplicity of light leptons and hadronically decaying τ (τ_{had}) candidates, are defined. They cover a wide range of yields and S/B purities, with different background composition in the various regions. The irreducible background, made of SM processes with prompt leptons, is dominated by $t\bar{t}W$, $t\bar{t}Z$ and diboson productions. It is estimated from NLO MC predictions and validated in data in various dedicated regions. The reducible background is mainly coming from $t\bar{t}$ events, with non-prompt light leptons from b -hadron decay, fake τ_{had} candidates from light flavour jet or charge misidentified electrons. These contributions are mostly estimated through data-driven techniques,

and validated in various regions. Dedicated MVA approaches are used in most SR to further enhance the signal to background separation. A combined profile likelihood fit across all SR and CR extracts the signal with an observed (expected) significance of 4.1σ (2.8σ). The measured signal strength is in good agreement with the SM prediction within the relative measurement precision of 30%, as seen in Figure 1 (right). This precision is fairly shared between statistical and systematic errors, dominated by uncertainties on $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) modelling and those associated to non-prompt lepton estimates (with a large contribution from the limited CR statistics). Similar results are obtained with an alternative fit using a free floating $t\bar{t}V$ normalisation, albeit with a 15% degraded sensitivity.

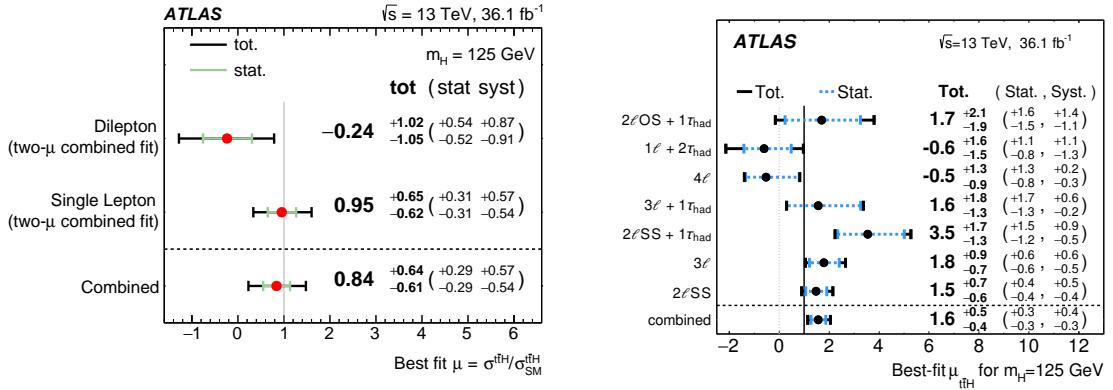


Figure 1: Observed best-fit values of the $t\bar{t}H$ signal strength μ and their uncertainties by final-state category and combined, for the $t\bar{t}H$, $H \rightarrow b\bar{b}$ (left [2]) and multileptons (right [3]) analyses. The individual μ values for the channels are obtained from a simultaneous fit with the signal-strength parameter for each channel floating independently. The SM prediction is $\mu = 1$.

4 Measurement in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$ ($l = e, \mu$) channels

The $t\bar{t}H$, $H \rightarrow \gamma\gamma$ channel [4] is more rare, but offers a clean signature with a narrow peak in the diphoton invariant mass distribution over a smooth continuum background, extrapolated from sideband regions. Two high- E_T isolated photons are required, and events are categorised in leptonic and hadronic regions depending on the presence of a light lepton. A further signal to background separation power is based on BDT discriminants, trained with deep neural networks using $t\bar{t}H$ MC and background events from data control regions. Seven categories (4 hadronic and 3 leptonic) are defined according to the value of the BDT output. The signal over background

ratio is ~ 1.9 in the purest categories, with the background dominated by the diphoton continuum (including $t\bar{t}+\gamma\gamma$). A combined unbinned fit to the diphoton invariant mass distributions in the 7 categories is performed (see Figure 2 (left)), with the Higgs signal modelled with a double-sided Crystal Ball function, and the continuum background with a functional form whose parameters are free in the fit. The significance of the observed (expected) signal is 4.1σ (3.7σ). The measured signal strength is in good agreement with the SM prediction within the relative measurement precision of 30%, dominated by statistical errors. The dominant systematic uncertainties are associated to the signal parton shower model (8%) and the photon energy resolution (6%).

A similar strategy is performed in the $t\bar{t}H$, $H \rightarrow ZZ^* \rightarrow 4l$ ($l = e, \mu$) channel [4], with a categorisation in leptonic and hadronic regions and the use of a BDT discriminant in the hadronic category. The signal over background ratio is ≥ 5 in the purest categories, with the background dominated by $t\bar{t}V$ and non- $t\bar{t}H$ Higgs production modes. The expected significance is 1.2σ (0.6 $t\bar{t}H$ events), while 0 event are observed.

5 Combination and results

The combination [4] of the searches for $t\bar{t}H$ production described in the previous sections is based on a simultaneous fit to all signal and control regions of the individual analyses. The contributions from the non- $t\bar{t}H$ Higgs production modes are fixed to SM predictions. Using $\sqrt{s} = 13$ TeV data only, an observed (expected) signal significance of 5.8σ (4.9σ) is achieved. Combining with $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV data increases the observed (expected) signal significance to 6.3σ (5.1σ). This constitutes an observation of the $t\bar{t}H$ process, which was a major milestone for LHC Run-2. The total $t\bar{t}H$ cross-section at $\sqrt{s} = 13$ TeV is measured to be 670 ± 90 (stat) $^{+110}_{-100}$ (syst) fb, in good agreement with the NLO SM prediction of 507^{+35}_{-50} fb. The measurement precision of 20% is slightly dominated by systematic uncertainties, in particular from $t\bar{t}$ +heavy flavour modelling (9.9%) and signal modelling (6.0%). The cross-section measurements performed at $\sqrt{s} = 13$ TeV and $\sqrt{s} = 8$ TeV are shown in Figure 2 (right).

Measurements of Higgs boson couplings combining several ATLAS analyses, including those on $t\bar{t}H$ reported here, have been performed [5]. Assuming SM particles only ($gg \rightarrow H$ and $H \rightarrow \gamma\gamma$ loops are then resolved with SM particles), the top quark Yukawa coupling, relative to its SM value, is measured to be $\kappa_{top} = 1.03^{+0.12}_{-0.11}$. Allowing new particles to enter in the loops, an effective Higgs-to-gluon coupling is defined, and a ratio $\kappa_{gluon}/\kappa_{top} = 1.09 \pm 0.14$ is extracted. This is consistent with SM Higgs boson coupling predictions, and allows to put constraints on BSM contributions.

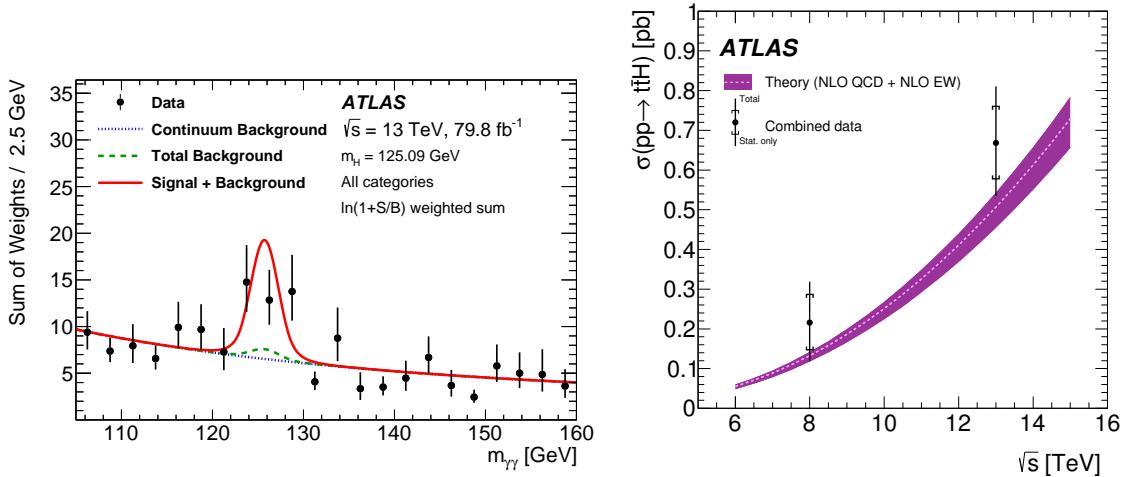


Figure 2: Left: Weighted diphoton invariant mass spectrum in the $t\bar{t}H$ -sensitive BDT bins. The solid red curve shows the fitted signal-plus-background model. The non-resonant and total background components of the fit are shown with the dotted blue curve and dashed green curve. Right: Measured $t\bar{t}H$ cross-sections in pp collisions at center-of-mass energies of 8 TeV and 13 TeV. Both the total and statistical-only uncertainties are shown. The measurements are compared with the SM prediction. Ref. [4].

6 Conclusions

The $t\bar{t}H$ production process has been observed using pp collision data produced by the LHC and recorded with the ATLAS detector. The cross-section at $\sqrt{s} = 13$ TeV is measured with 20% precision. All measurements are in good agreement with SM predictions, allowing to put constraints on BSM contributions. This constitutes a direct observation of the top quark Yukawa coupling, which, together with the recent observations of $H \rightarrow \tau\tau$ and $H \rightarrow b\bar{b}$ decays by the ATLAS and CMS collaborations separately, firmly establishes the existence of Yukawa interactions.

References

- [1] ATLAS Collaboration, JINST **3** (2008) S08003.
- [2] ATLAS Collaboration, Phys. Rev. D **97** (2018) 072016.
- [3] ATLAS Collaboration, Phys. Rev. D **97** (2018) 072003.
- [4] ATLAS Collaboration, Phys. Lett. B **784** (2018) 173.
- [5] ATLAS Collaboration, ATLAS-CONF-2018-031.